

---

Lina Merchan · Ilya Nemenman

# On the sufficiency of pairwise interactions in maximum entropy models of biological networks

May 13, 2015

**Abstract** Biological information processing networks consist of many components, which are coupled by an even larger number of complex multivariate interactions. However, analyses of data sets from fields as diverse as neuroscience, molecular biology, and behavior have reported that observed statistics of states of some biological networks can be approximated well by maximum entropy models with only pairwise interactions among the components. Based on simulations of random Ising spin networks with  $p$ -spin ( $p > 2$ ) interactions, here we argue that this reduction in complexity can be thought of as a natural property of densely interacting networks in certain regimes, and not necessarily as a special property of living systems. By connecting our analysis to the theory of random constraint satisfaction problems, we suggest a reason for why some biological systems may operate in this regime.

**Keywords** collective dynamics ·  $p$ -spin models · numerical simulations

## 1 Introduction

The increased throughput of biological experiments now allows joint measurements of activities of many basic components underlying collective information processing in biological systems. Such multivariate data must be interpreted within models. Within this context, Maximum Entropy (MaxEnt) models [1] have been some of the most successful. The logic of such models is that, ultimately, one wants to find an approximation  $Q(\mathbf{x})$  to the joint probability distribution  $P(\mathbf{x})$  of the observed multivariate data  $\{x_i\} = \mathbf{x}$ ,

---

Lina Merchan

Department of Physics, Emory University, Atlanta, GA 30322, USA

Ilya Nemenman E-mail: [ilya.nemenman@emory.edu](mailto:ilya.nemenman@emory.edu)

Departments of Physics and Biology, Emory University, Atlanta, GA 30322, USA

$i = 1, \dots, N$ . Unfortunately, for a large number of components,  $N$ , the datasets can never be large enough to estimate  $P$  directly from data. One may only be able to estimate various expectation values of functions of the data,  $\langle f_\kappa(\mathbf{x}) \rangle_P = \tilde{f}_\kappa$ ,  $\kappa = 1, \dots, K$ . Then one can search for  $Q$  that matches the reliable estimates. If additionally one requests that  $Q$  has no structure beyond that required by the matching, then this is equivalent to asking for  $Q$  with the maximum entropy, subject to the constraints imposed by the matching,

$$Q = \arg \max S(Q) - \sum_{\kappa} \lambda_{\kappa} (\langle f_{\kappa} \rangle_Q - \tilde{f}_{\kappa}), \quad (1)$$

where the entropy  $S$  is defined as

$$S(Q) \equiv S(\mathbf{x}) = - \sum_{\mathbf{x}} Q(\mathbf{x}) \log_2 Q(\mathbf{x}). \quad (2)$$

A common special case of this general formulation is when the variables are binary, which we will denote as  $x_i = \sigma_i \in \{-1, 1\}$ , and the data constrain their various low-order correlation functions, such as  $\langle \sigma_i \rangle$  or  $\langle \sigma_i \sigma_j \rangle$ . In this case, the MaxEnt approximation  $Q$  is [2]:

$$Q(\boldsymbol{\sigma}) = \frac{1}{Z} \exp \left( - \sum_i h_i \sigma_i - \sum_{ij} J_{ij} \sigma_i \sigma_j - \sum_{ijk} K_{ijk} \sigma_i \sigma_j \sigma_k - \dots \right). \quad (3)$$

Here every constrained correlation function gets a term in the exponent,  $Z$  is the partition function, and the Lagrange multipliers  $h_i, J_{ij}, K_{ijk}, \dots$  must be chosen to satisfy the constraints. This is generally not an analytically solvable problem, and even numerics are hard [3, 4, 5, 6, 7, 8, 9].

Equation (3) has the form of the Ising spin problem, allowing a wholesale import of intuition from statistical physics to MaxEnt data analysis. Correspondingly, these ideas have been applied to many biological systems in the last decade [10], starting with neurophysiological recordings from salamander retina [11]. There  $N$  was a few dozen neurons, and  $\sigma_i = \pm 1$  corresponded to the  $i$ 'th neuron spiking/not spiking at a certain time. A surprising result was that truncating Eq. (3) at the quadratic order in  $\sigma_i$  (or, in other words, constraining  $Q$  up to pairwise correlations) provided a good fit to  $P$ . We will refer to this finding as *pairwise sufficiency* from now on.

The pairwise sufficiency was later found in other neural systems [12, 13, 14] (though it is violated at larger  $N$  [15]). It was observed further for natural images [16]; for discrete, yet non-binary  $\mathbf{x}$  in sequencing data [17, 18]; and for real-valued velocities of birds in flocking experiments [19]. Even for some non-MaxEnt approaches, similar findings were also reported [20, 21]. One can interpret these observations in the context of biological systems operating in a special regime [22, 23]. However, the wide applicability of the findings suggests an alternative: pairwise sufficiency may emerge for a wide class of biological and non-biological networks *generically*. Indeed, sparse sampling of variables in experiments is similar to decimation in statistical physics, and the resulting renormalization group-like flow may decrease the importance of the higher order couplings [24]. Further, in a perturbative regime, where

fluctuations away from the independence are small, the pairwise sufficiency also appears [25]. Here we propose one more possibility, arguing that the pairwise sufficiency arises naturally in *strongly coupled* multivariate systems.

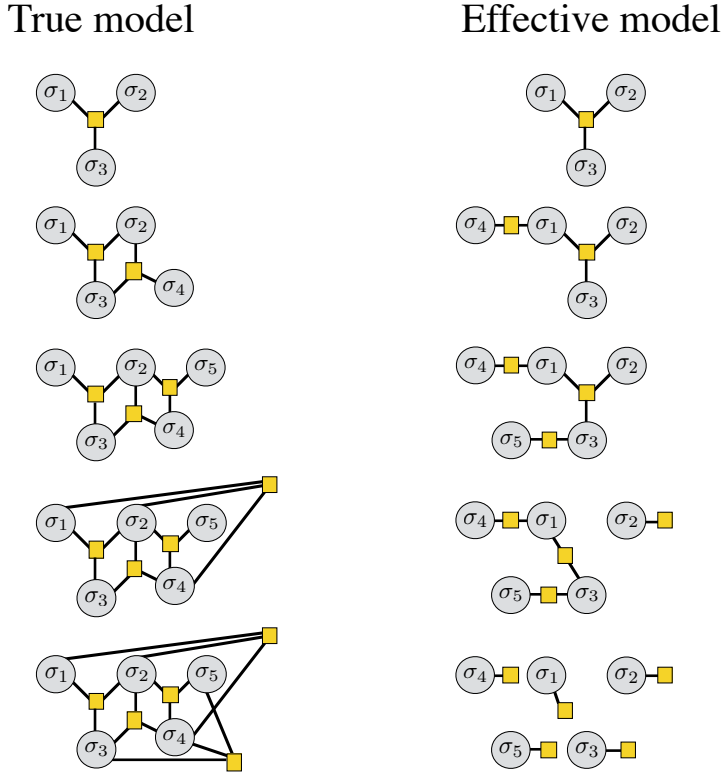
In what follows, we first introduce the idea in an intuitive toy model, and then develop it numerically by analyzing randomly generated networks. We show that the pairwise MaxEnt models approximate such random networks surprisingly well. Further, we explore distributions of states of these networks and their models, leading to an explanation of the pairwise sufficiency. Finally, we discuss why diverse biological systems may find themselves in the pairwise-sufficient regime, but we leave it for the future to investigate if this mechanism is responsible for the sufficiency in experimental networks.

## 2 Results

### 2.1 Building Intuition: Networks of XORs

For a tractable example of emergence of the pairwise sufficiency, we focus on Boolean gates. These are the limit of Ising spin networks in the low temperature (strong coupling) regime [2, 26]. For example,  $\sigma_3 = \sigma_1 \text{ OR } \sigma_2$  can be written as  $P(\sigma_3|\sigma_2, \sigma_1) = \frac{1}{2} \exp[J(\sigma_1\sigma_3 + \sigma_2\sigma_3 + \sigma_3)]$  with  $J \rightarrow \infty$ . If also  $P(\sigma_1 = \pm 1) = P(\sigma_2 = \pm 1) = 1/2$ , then  $1/4 P(\sigma_3|\sigma_2, \sigma_1) = P(\sigma_1, \sigma_2, \sigma_3)$ . Thus the joint probability distribution for OR has the pairwise MaxEnt form, Eq. (3). Similarly, for  $J \rightarrow \infty$ ,  $\sigma_3 = \sigma_1 \text{ AND } \sigma_2$  is equivalent to  $P(\sigma_1, \sigma_2, \sigma_3) = \frac{1}{2} \exp[J(\sigma_1\sigma_3 + \sigma_2\sigma_3 - \sigma_3)]$ . This is again a pairwise MaxEnt distribution. However,  $\sigma_3 = \sigma_1 \text{ XOR } \sigma_2$ , is equivalent to  $P(\sigma_1, \sigma_2, \sigma_3) = \frac{1}{2} \exp(-K\sigma_1\sigma_2\sigma_3)$ ,  $K \rightarrow \infty$ . This is an example of a purely third-order gate, with no pairwise contributions to its MaxEnt representation.

In Fig 1, we now couple a few such third-order gates to each other. The spins  $\sigma_1, \sigma_2, \sigma_3$  are connected by an XOR (left column, first row), and there is no simpler effective representation of the network (right column, first row). We then add the fourth spin,  $\sigma_4 = \sigma_2 \text{ XOR } \sigma_3$  (left column, second row). However, then  $\sigma_4 = \sigma_1$ . This can be represented as an effective model  $P(\sigma_4|\sigma_1, \dots, \sigma_3) = P(\sigma_4|\sigma_1) = \frac{1}{2} \exp(J\sigma_1\sigma_4)$ ,  $J \rightarrow \infty$ . Thus the third order XOR interaction is equivalent to a pairwise EQUAL interaction (right column, second row). The latter is effective and nonlocal, in the sense that  $\sigma_4$  is coupled to  $\sigma_1$ , with which it does not interact in the true model. We further add  $\sigma_5 = \sigma_2 \text{ XOR } \sigma_4$  (third row), and this is equivalent to an effective model  $\sigma_5 = \sigma_3$ . In short, of the three third order interactions, each constraining one spin and hence “carrying” 1 bit of information, two can be represented without any error as pairwise interactions. Now the network can exist in four distinct global states out of  $2^5 = 32$ , determined by  $\sigma_{1,2} = \pm 1$  (namely,  $\boldsymbol{\sigma}^{(1)} = \{-1, -1, -1, -1, -1\}$ ,  $\boldsymbol{\sigma}^{(2)} = \{-1, +1, +1, -1, +1\}$ ,  $\boldsymbol{\sigma}^{(3)} = \{+1, -1, +1, +1, +1\}$ , and  $\boldsymbol{\sigma}^{(4)} = \{+1, +1, -1, +1, -1\}$ ). Thus it is far from the perturbative regime of Ref. [25]. We can grow the network further so that each new spin is coupled by a third order interaction to two existing spins. Then the number of spins,  $N$ , and the number of interactions,  $M$ , are related as  $N = M + 2$ , and all but one third order interaction can be represented as



**Fig. 1 Emergence of pairwise interactions in a network of XOR gates.** On the left, we show small networks of spins  $\sigma_i$  (grey circles). The spins interact (yellow squares) by means of third order XOR interactions. On the right, an equivalent network is shown, where some of the XORs get replaced by EQUAL and assignment operations, which are the second and the first order interactions, respectively.

a second order one. In other words, an effective pairwise model has an error of only  $1/(N-2)$  when accounting for the statistics of the network states.

Alternatively, we can add more XORs without adding new nodes. This may be inconsistent or redundant with already existing couplings. Or in a case such as  $\sigma_1 = \sigma_2 \text{ XOR } \sigma_4$  (fourth row), this sets  $\sigma_2 = -1$  (thus adding the bias, or the first order term), and all other spins are equal to each other, so that the pairwise effective model is exact. Finally, adding  $\sigma_3 = \sigma_4 \text{ XOR } \sigma_5$  sets every spin to -1, and makes even the first-order model exact (bottom row).

We see that a network of XORs can exhibit the pairwise sufficiency non-perturbatively. Of course, more realistic physical or biological systems are stochastic ( $J, K < \infty$ ), and such simple arguments will not work. However, the example suggests that effective pairwise models can approximate more complex networks well when nodes in the network interact strongly and densely, and the space of network states is sufficiently constrained. In such cases, there are many pairs of nodes that are relatively strongly correlated simply by chance, allowing replacement of higher order interactions

with pairwise ones. In what follows, we will develop this intuition further using numerical studies.

## 2.2 Pairwise approximations to random networks with higher order interactions

To verify our intuition, we proceed by generating random networks that have *only* higher order nondeterministic interactions among spins ( $p$ -spin models,  $p > 2$  [27]). We then quantify the accuracy with which lower order MaxEnt models approximate these networks. We explore networks with  $p = 3, 4$  to ensure that our findings do not depend on the exact structure of the true higher order interactions. Further, systems with only fourth order couplings have the  $Z_2$  symmetry, and thus cannot include any first order terms in their MaxEnt approximations, Eq. (3). Studying them will allow us to understand if the eventual freezing to a single well-defined state, as in the last row of Fig. 1, is crucial for the pairwise sufficiency, or if it emerges even for nonperturbative networks with more than one highly probable state.

To generate the random networks, we first specify  $N$ , the number of nodes, and  $M$  the number of interactions. Then for each interaction  $\mu = 1, \dots, M$ , we generate its coupling constant  $K_\mu$  from a zero-mean Gaussian distribution with a certain variance  $s^2$ . We then choose three or four spins at random to couple. The overall probability of states for these networks is

$$P_3(\boldsymbol{\sigma}) = \frac{1}{Z} \exp \left( - \sum_{\mu=1}^M K_\mu \sigma_{\mu_1} \sigma_{\mu_2} \sigma_{\mu_3} \right), \text{ 3-spin model}, \quad (4)$$

$$P_4(\boldsymbol{\sigma}) = \frac{1}{Z} \exp \left( - \sum_{\mu=1}^M K_\mu \sigma_{\mu_1} \sigma_{\mu_2} \sigma_{\mu_3} \sigma_{\mu_4} \right), \text{ 4-spin model}, \quad (5)$$

where  $\mu_1 < \mu_2 < \mu_3 < \mu_4$ , so that the spins do not self-interact. To specify these distributions (and later calculate various errors of approximations), we need to know  $Z$ . To decouple studying the problem of the pairwise sufficiency from a hard problem of efficient sampling, we focus on  $N \leq 22$ , which allows us to estimate  $Z$  by direct summation fast enough to do it many times and collect statistics. We generate many such distributions  $P_3$  and  $P_4$ , every time picking random  $N \in [10, 22]$ ,  $M \in [1, 250]$ , and  $s \in [0.2, 2.0]$ .

For each generated distribution, we estimate its individual and pairwise marginals  $P(\sigma_i)$ ,  $P(\sigma_i, \sigma_j)$  for all  $i, j = 1, \dots, N$  by direct marginalization (hereafter we drop subscripts 3 or 4 for  $P$  if it does not cause confusion). We then calculate the first order (or independent) MaxEnt approximation

$$Q^{(1)}(\boldsymbol{\sigma}) = \prod_{i=1}^N P(\sigma_i). \quad (6)$$

Next we fit the pairwise MaxEnt model  $Q^{(2)}$  to  $P$ . While good algorithms exist for this purpose [3, 4, 5, 6, 7, 8, 9], it is unclear if their assumptions are satisfied by our networks. Trying again to decouple the problems of efficient

inference and the pairwise sufficiency, we choose a classic, well understood Iterative Proportional Fitting Procedure (IPFP) algorithm [28]. That is, we start with  $Q^{(1)}$  as a guess for  $Q^{(2)}$ , calculate  $Q^{(2)}(\sigma_i, \sigma_j)$ , and redefine

$$Q^{(2)}(\boldsymbol{\sigma}) \rightarrow Q^{(2)}(\boldsymbol{\sigma}) \frac{P(\sigma_i, \sigma_j)}{Q^{(2)}(\sigma_i, \sigma_j)}. \quad (7)$$

We cycle through all pairs  $i, j$ , and iterate until  $Q^{(2)}(\sigma_i, \sigma_j) \approx P(\sigma_i, \sigma_j)$  up to a relative error of  $10^{-5}$ . This is achieved within  $\sim 10^0 \dots 10^4$  iterations depending on how close the final  $Q^{(2)}$  is to  $Q^{(1)}$ . We verified that starting with different initial conditions results in the same solution, as it should.

To measure the quality of the MaxEnt models, we calculate the Kullback-Leibler (KL) divergence between the true distribution  $P_{3,4}$  and each approximation, normalized by the number of spins in the system:

$$\mathcal{D}^{(1)} = \frac{D_{\text{KL}}^{(1)}}{N} \equiv \frac{1}{N} \sum_{\boldsymbol{\sigma}} P(\boldsymbol{\sigma}) \log_2 \frac{P(\boldsymbol{\sigma})}{Q^{(1)}(\boldsymbol{\sigma})}, \quad (8)$$

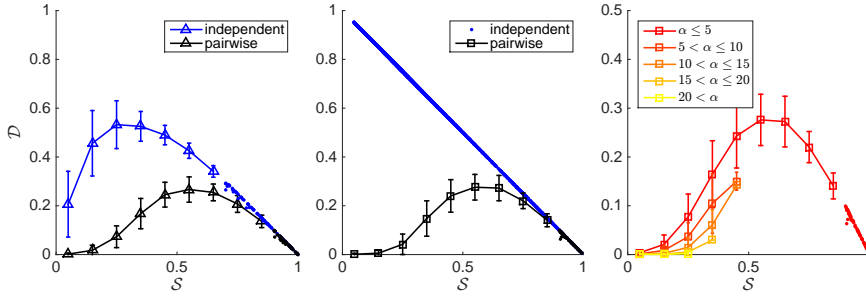
$$\mathcal{D}^{(2)} = \frac{D_{\text{KL}}^{(2)}}{N} \equiv \frac{1}{N} \sum_{\boldsymbol{\sigma}} P(\boldsymbol{\sigma}) \log_2 \frac{P(\boldsymbol{\sigma})}{Q^{(2)}(\boldsymbol{\sigma})}. \quad (9)$$

Since our maximum  $N$  is rather small, this is done by direct summation. Notice that both  $\mathcal{D}^{(1)}$  and  $\mathcal{D}^{(2)}$  are between zero (perfect fit) and one (the worst fit) if single-spin marginals of  $Q$  and  $P$  are equal.

In Fig. 2, we plot the values of  $\mathcal{D}^{(1)}$  and  $\mathcal{D}^{(2)}$  measured over different ensembles of random networks vs. the normalized entropy of the network's state space  $\mathcal{S} = S(\boldsymbol{\sigma})/N$ , which also varies between 0 and 1. For all types of networks and approximation, the quality of fit is high ( $\mathcal{D}^{(\cdot)}$  is low) when  $\mathcal{S} \sim 1$ , so that the networks are unconstrained, and nearly all states are possible. This is trivial since even the zeroth order approximation (each spin up or down with 50% probability) would work well here.

As  $\mathcal{S}$  decreases, the fit errors increase. When  $\mathcal{S}$  reaches small values, the independent approximation,  $\mathcal{D}^{(1)}$ , starts behaving differently for the different network types. In the 4-spin case, by construction,  $P(\boldsymbol{\sigma}) = P(-\boldsymbol{\sigma})$ . Thus  $\langle \sigma_i \rangle = 0$  for any  $i$ , and the best independent approximation is the uniform distribution. For this construction, the smallest possible entropy is  $\mathcal{S} = 1/N$ , where the network exists in two mirror states, and there the error of  $Q^{(1)}$  is  $\mathcal{D}_4^{(1)} = 1 - 1/N$ . In contrast, a 3-spin network freezes at  $\mathcal{S} = 0$ , and each spin is strongly biased (as in our XOR networks above). Thus the independent approximation provides a perfect fit in this case.

The distinction between  $P_3$  and  $P_4$  vanishes for the pairwise MaxEnt approximation. Here, for both 3- and 4-spin networks, the fit errors behave similarly: for  $\mathcal{S}$  decreasing from 1,  $\mathcal{D}^{(2)}$  grows from 0 and reaches its peak at about  $\mathcal{D}^{(2)} \approx 0.25 \dots 0.3$  near  $\mathcal{S} \approx 0.5 \dots 0.6$ . This is already interesting:  $\mathcal{D}^{(2)}$  almost never goes above 0.3 for all networks we tried. Thus even in some of the worst cases, pairwise approximation is quite good! Further, for even smaller entropies,  $\mathcal{D}^{(2)}$  rapidly drops, approaching zero faster than linearly in  $\mathcal{S}$ . For  $\mathcal{S} \approx 0.25$ ,  $\mathcal{D}_3^{(2)} \approx 0.07$ . It is even smaller,  $\mathcal{D}_4^{(2)} \approx 0.04$ , for the quartic

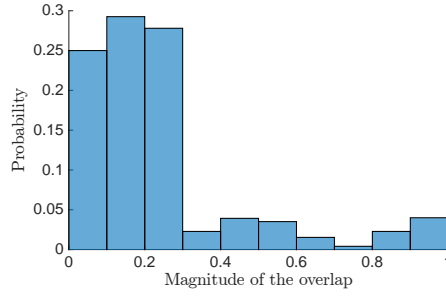


**Fig. 2 Error of the MaxEnt fits vs. the normalized entropy of the network state space,  $\mathcal{S}$ .** The left panel shows errors of the independent,  $\mathcal{D}^{(1)}$ , and the pairwise,  $\mathcal{D}^{(2)}$ , approximations for 3-spin networks. We used  $\sim 800$  random networks with  $11 \leq N \leq 20$  spins and with a varying number of interactions,  $M$ . We partitioned all the networks by their  $\mathcal{S}$  in bins of width of 0.1 and calculated the mean and the standard deviation of  $\mathcal{D}$  for each bin. These are indicated by triangles and the error bars. Wherever the data points for individual networks showed little scatter, we plotted these points instead of the bin averages. The middle panel presents similar data,  $\mathcal{D}^{(1)}$  and  $\mathcal{D}^{(2)}$ , for 4-spin networks. Here over 4000 random networks were generated with  $11 \leq N \leq 22$ .  $\mathcal{D}^{(2)}$  was again averaged within ten bins, and the means and the standard deviations are plotted. For  $\mathcal{D}^{(1)}$ , data for individual networks are presented. These merge into a perfect straight line due to the  $Z_2$  symmetry of 4-spin distributions. For both the 3- and the 4-spin cases, the pairwise sufficiency is clear at low  $\mathcal{S}$ . The right panel replots the  $\mathcal{D}^{(2)}$  data for the 4-spin networks, but splits them according to  $\alpha$ , which measures the average strength of interactions per spin within a network. Large  $\alpha$  curves are significantly below their small  $\alpha$  counterparts, indicating that, other things being equal, densely and strongly interacting networks are more likely to be pairwise sufficient. Notice that large  $\alpha$  curves end abruptly since such networks cannot have large  $\mathcal{S}$ .

case. This is because  $\min \mathcal{S}_4 = 1/N$ , so that the whole  $\mathcal{D}_4^{(2)}$  curve is slightly shifted compared to  $\mathcal{D}_3^{(2)}$  at low  $\mathcal{S}$ .

In summary, for all the networks we have considered, pairwise sufficiency emerges robustly at low (but not too low) entropy. In fact, at  $\mathcal{S} \approx 0.25$ , our networks can be in more than  $2^{\mathcal{S}} = 2^{N\mathcal{S}} \approx 2^5 = 32$  highly probable states. Thus the networks are not totally frozen, and yet the pairwise approximation is nearly sufficient! Crucially, this finding is robust to the changes in the network size:  $\mathcal{D}$  vs.  $\mathcal{S}$  curves are stable over the entire range of  $N$  we explored.

Within a single narrow bin of  $\mathcal{S}$ ,  $\mathcal{D}^{(2)}$  may still have a rather large range. We explore this variability in the rightmost panel of Fig. 2. For this, we define  $\alpha = sM/N$  (recall that  $s$  is the standard deviation of the random couplings used to generate the networks).  $\alpha$  measures the strength of interactions (or constraints) per spin, analogously to a similar parameter in the random constraint satisfaction problems [29,30]. For quartic networks, where we have enough samples, we then plot  $\mathcal{D}^{(2)}$  vs.  $\mathcal{S}$  for different ranges of  $\alpha$ . Crucially, we find that, for the same  $\mathcal{S}$ , a larger  $\alpha$  results in better pairwise fits. In other words, a denser and stronger interacting network is more likely to be pairwise sufficient. This is potentially a good news for MaxEnt approaches



**Fig. 3 The pairwise sufficiency emerges in the nonperturbative regime.**

We select all highly probable states  $\sigma^\mu$ , defined somewhat arbitrary as  $P(\sigma^\mu) > 0.001$ . We then calculate the overlap for all pairs of such states. Finally, we plot the histogram of the magnitudes of the overlaps from all 4-spin networks with  $N \geq 20$ , and with  $0.1 \leq \mathcal{S} \leq 0.3$ . Such networks can exist in many states, but still very few compared to  $2^N$ . Most of the overlap magnitudes are away from 1, indicating small to moderate similarity among the highly probable states. Thus these states are broadly dispersed and do not cluster together.

to biological systems, which are known for the immense complexity of the underlying biophysical interactions.

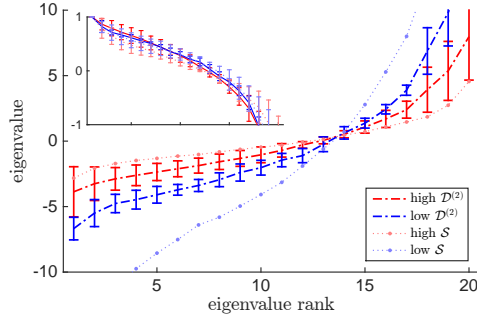
We conclude this section by stressing that high probability states of pairwise sufficient  $p$ -spin networks are not close to each other. To illustrate this, we focus on the 4-spin case with  $N \geq 20$ , and on small but not negligible  $\mathcal{S}$ . We then evaluate the magnitude of the overlap,  $|\sigma^\mu \cdot \sigma^\nu|/N$ , among all highly probable network states and plot the distribution of the overlaps in Fig. 3. For purely randomly distributed states, we would expect the standard deviation of overlaps to be  $\sim 0.22$ , and a peak near zero. And we would expect magnitudes of overlaps near 1 if all highly probable states were clustered near a dominant one. Instead, the distribution in Fig. 3 is not concentrated near 1, and the standard deviation is  $\approx 0.39$ . Therefore, there is some clustering of probable states, but certainly not strong clustering. Thus the state space of our networks cannot be described as small fluctuations around a dominant state, and the pairwise sufficiency here is not perturbative [25]. It likely emerges due to a previously not investigated mechanism.

### 2.3 The structure of the state space of the pairwise sufficient networks

The toy example of the XOR network suggests that the pairwise sufficiency may emerge when the network “freezes” to a few (but not necessarily just one) highly probable states, and different relatively tightly coupled clusters of spins decouple from each other. Is this also true for our networks with a nonzero temperature? How do energy landscapes of the sufficient and the insufficient networks differ from each other? And are the MaxEnt fits for both cases structurally different?

To start exploring this, we estimate  $h_i$  and  $J_{ij}$  for  $Q^{(2)}$  inferred using IPFP. We do this by choosing  $\gg N(N+1)/2$  states with the highest probability from  $Q^{(2)}$ . We get the energy of each such state as  $E(\sigma) = -\log P(\sigma)$

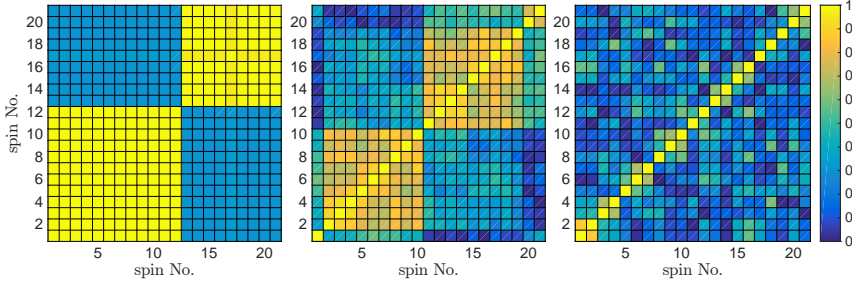




**Fig. 4 Spectra of the pairwise coupling matrices  $J_{ij}$  for MaxEnt approximations to random 4-spin networks,  $N \geq 20$ .** We order eigenvalues from the smallest (lowest energy) to the largest (highest energy). We then plot the mean spectra (with standard deviations, where it does not obstruct the figures), averaged over different subsets of 4000 networks. Low  $\mathcal{S}$  subset corresponds to  $0.1 \leq \mathcal{S} < 0.2$ . Such networks are fit extremely well by pairwise models, with mean  $\mathcal{D}^{(2)} \approx 6 \cdot 10^{-3}$ . High  $\mathcal{S}$  corresponds to  $0.3 \leq \mathcal{S} < 0.4$ . Here the pairwise fits are bad, so that the mean  $\mathcal{D}^{(2)} \approx 0.15$ . Finally, for the intermediate range  $0.2 \leq \mathcal{S} < 0.3$ , the quality of fits is diverse. We further partition this range into well fitted,  $\mathcal{D}^{(2)} \leq 0.06$ , and badly fitted,  $\mathcal{D}^{(2)} \geq 0.12$ , subsets, leaving intermediate fits off the plot. The four average spectra show that the pairwise sufficiency is directly correlated with the scale of the spectra, with larger magnitude eigenvalues resulting in smaller  $\mathcal{D}^{(2)}$ . The inset shows the averages for each of the four ranges, where each spectrum is normalized by its largest magnitude negative eigenvalue. The four curves are very close for much of their range.

and then solve the linear regression problem to find the coupling constants from the states and their energies. Finally, we calculate the eigenvalue spectra of the inferred  $J_{ij}$ , having set  $J_{ii} = 0$ . The averaged spectra are shown in Fig. 4 for different combinations of  $\mathcal{S}$  and  $\mathcal{D}^{(2)}$ . We see that the success of  $Q^{(2)}$  is correlated with the magnitude of the eigenvalues of  $J_{ij}$  — larger magnitudes, which correspond to stronger interactions and more constrained distributions, give the pairwise sufficiency. This is true irrespective of  $\mathcal{S}$  (though  $\mathcal{S}$  and  $\alpha$  are dependent, as we have discussed). Crucially, if one rescales the spectra by their largest magnitude negative eigenvalue (Fig. 4, inset), then all spectra collapse. Thus the  $J_{ij}$  (or the energy landscapes) of the pairwise sufficient and the pairwise insufficient fits are not intrinsically different: a rescaling (change in temperature) can morph one into the other.

Having analyzed the pairwise sufficient and insufficient solutions,  $Q^{(2)}$ , we now focus on the landscapes of the  $p$ -spin networks themselves. The freezing that results in the decrease of  $\mathcal{D}^{(2)}$  and the growth of the eigenvalues of  $J$  can create the landscapes of different types. For example, the highly probable states may be essentially uncorrelated, reminiscent of the landscapes of the Hopfield network in the ferromagnetic phase [31]. Alternatively, as in our XOR network, entire blocks of spins can merge into strongly correlated clusters, which then decouple from each other. Then the low energy network states will be direct products of the states of the clusters. To disambiguate the two scenarios, we calculate pairwise spin-spin correlation  $c_{ij} = \frac{\text{cov}(\sigma_i, \sigma_j)}{\text{std } \sigma_i \text{std } \sigma_j}$



**Fig. 5 4-spin networks decouple into spin clusters in the pairwise sufficient regime.** The three panels show typical values of  $|c_{ij}|$  for  $N = 21$ , sorted into correlated clusters. Left panel: a nearly perfectly pairwise sufficient network with  $S = 2.05$  bits and  $\mathcal{D}^{(2)} = 8.0 \cdot 10^{-5}$ . In accord with the entropy value, the network splits into two clusters, with spins nearly perfectly correlated within each, but almost independent across. Middle panel: a good pairwise fit with  $S = 3.9$  bits and  $\mathcal{D}^{(2)} = 0.032$ . Correspondingly, four clusters of different sizes are seen. However, now the spins also exhibit some correlations across clusters, which presumably leads to the increase in  $\mathcal{D}^{(2)}$ . Right panel: a network with  $S = 5.7$  bits and  $\mathcal{D}^{(2)} = 0.16$  – a bad (though not disastrous) pairwise fit. There are now many small clusters, but correlations within and across the clusters are not very different.

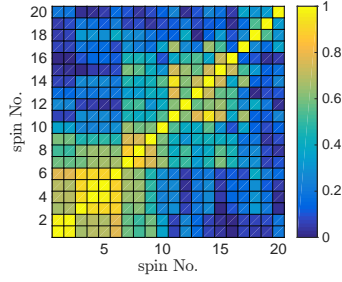
by direct summation of  $P$  (note that, for 4-spin networks, the correlation is equal to the covariance since  $\langle \sigma_i \rangle = 0$ ). We then cluster the spins based on the absolute value of their correlations. Figure 5 shows the clusters for 4-spin networks (note that since the number, the size, and the spin assignment for clusters are different for each network, we only show typical cases). A network with a near-zero  $\mathcal{D}^{(2)}$  (left panel) shows a perfect partitioning into two clusters;  $S(\boldsymbol{\sigma}) \approx 2$  bits is a result of this partitioning. As networks with larger entropies are considered, the number of clusters increases, and their boundaries become fuzzy, leading to worse MaxEnt fits. When the definite cluster structure disappears,  $\mathcal{D}^{(2)}$  grows dramatically. Thus the existence of well-defined spin clusters is correlated with the pairwise sufficiency.

For 3-spin networks, in addition to the pairwise interactions, there are also nonzero single spin biases in the MaxEnt fits. Thus the entropy and correlations among spins are generally smaller for the same  $\mathcal{D}^{(2)}$ . Nonetheless, as seen in Fig. 6, the (fuzzy) cluster structure for these networks is not that much different from the 4-spin case.

To further explore the network landscapes, we point out that an inferred symmetric  $J_{ij}$  can be rewritten as

$$J_{ij} = \sum_{\nu=1}^N \lambda^{(\nu)} \xi_i^{(\nu)} \xi_j^{(\nu)}, \quad (10)$$

where  $\lambda^{(\nu)}$  and  $\xi_i^{(\nu)}$  are the eigenvalues and the eigenvectors, correspondingly. The eigenvalues take both large positive and large negative values for the pairwise sufficient networks (cf. Fig. 4). The negatives correspond to wells in the landscape, and the positives correspond to peaks. If the wells and the peaks were clearly separated, then spin configurations in the vicinity of the



**Fig. 6 A 3-spin network also decouples into spin clusters in the pairwise sufficient regime.** For brevity, here we show  $|c_{ij}|$  only for one network similar to the middle panel in Fig. 5:  $S \approx 3.5$ ,  $\mathcal{D}^{(1)} \approx 0.58$ , and  $\mathcal{D}^{(2)} \approx 0.038$ . Four to seven partially overlapping clusters can be seen.

well center, distinct from it by just a single spin flip, would have similar high probabilities (this is what allows an eigenvector to act as a broad attractor in the Hopfield network [31]). The repulsive peaks far away from the wells would have little effect on  $\mathcal{D}^{(2)}$  since the probability of states away from the wells is small in the low temperature regime even without the peaks. In contrast, were spins to form tight clusters, flipping a single spin would not be allowed. Peaks would be needed to decrease the probability of such cluster-breaking states, and thus positive eigenvalues would affect  $\mathcal{D}^{(2)}$  strongly.

To verify which of the two scenarios holds, for 4-spin networks, we construct the coupling matrix and the pairwise MaxEnt distribution from only  $n \leq N$  eigenvalues,

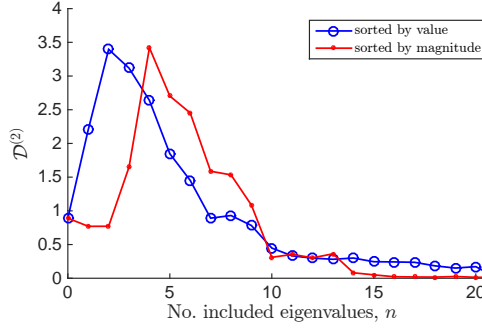
$$J_{ij}(n) = \sum_{\nu=1}^n \lambda^{(\nu)} \xi_i^{(\nu)} \xi_j^{(\nu)}, \quad (11)$$

$$Q^{(2)}(n) = \frac{1}{Z} \exp \left( - \sum_{i,j} J_{ij}(n) \sigma_i \sigma_j \right). \quad (12)$$

We then evaluate  $\mathcal{D}^{(2)}$  between  $P$  and  $Q^{(2)}(n)$  as a function of  $n$ . Figure 7 shows this dependence for a typical pairwise-sufficient distribution and for two different ways of including eigenvalues into  $J$ . In the first, we proceed from the most negative eigenvalue to the most positive one. In the second, we proceed from the largest magnitude eigenvalue to the smallest one. Since sorting by magnitude (which includes large positive eigenvalues earlier) approaches the terminal  $\mathcal{D}^{(2)}$  faster, wells and peaks must both affect close spin configurations. This is again consistent with the clustering picture.

#### 2.4 The mechanism of emergence of the pairwise sufficiency

The clustered structure of the network landscapes allows us to propose a hypothesis for why densely coupled  $p$ -spin networks exhibit the pairwise sufficiency. We re-group terms in the energies, which define  $P_3$  and  $P_4$  in



**Fig. 7 Positive eigenvalues of the MaxEnt coupling matrix  $J_{ij}$  contribute to the approximation error.** We plot the fit error,  $\mathcal{D}^{(2)}$ , as a function of the number of the eigenvalues of  $J_{ij}$  included in the fit for a 4-spin distribution with  $N = 21$ ,  $\mathcal{S} \approx 0.1$ , and  $\mathcal{D}^{(2)} \approx 0.014$ . The blue line includes eigenvalues in the order from the most negative to the most positive, and the red one includes them in the order of their absolute values. The red line reaches the limiting value of  $\mathcal{D}^{(2)}$  quicker, while the blue one requires inclusion of all eigenvalues for this to happen. As explained in the text, this is a signature of emergence of spin clusters.

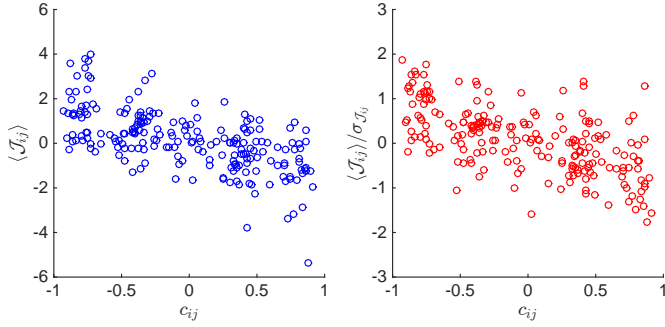
Eqs. (4, 5). For example, all terms that couple  $\sigma_i$  and  $\sigma_j$  for the 4-spin network can be rewritten as

$$\sigma_i \sigma_j \sum_{\mu} K_{\mu} \sigma_{\mu_3} \sigma_{\mu_4} \delta_{i, \mu_1} \delta_{j, \mu_2} \equiv \sigma_i \sigma_j \mathcal{J}_{ij}, \quad (13)$$

where  $\delta_{\cdot, \cdot}$  is a Kronecker delta. (Here we slightly abused the notation and imposed that  $i, j$  only occur in  $\mu_1$  and  $\mu_2$ .) This equation defines a random coupling  $\mathcal{J}_{ij}$ , which depends both on the current network state and on the quenched randomness that went into building the network. For a large number of couplings, fluctuations in  $\mathcal{J}_{ij}$  will be large enough so that, averaged over the accessible network states,  $\mathcal{J}_{ij}$  stays far from zero compared to its standard deviation  $\sigma_{\mathcal{J}_{ij}}$ . This creates large effective pairwise coupling among spins, so that clusters of spins start behaving coherently. Then the state of every spin in the cluster can be defined by choosing a cluster representative, setting its value, and then coupling each cluster member to the representative through a pairwise interaction — higher order couplings are not needed! The pairwise MaxEnt fit is nearly exact, even though the network is far from frozen since values of the cluster representatives are not necessarily constrained. We illustrate this in Fig. 8, which shows that higher order couplings average to produce large effective pairwise interactions for correlated spins.

### 3 Discussion

In this numerical study, we showed that pairwise MaxEnt models are more effective in approximating random  $p$ -spin networks ( $p = 3, 4$ ) than one would naively expect. Even in the worst cases, the error of such models was rarely



**Fig. 8 Formation of effective couplings in 4-spin networks.** For the same network as in Fig. 5, middle, we plot  $\langle \mathcal{J}_{ij} \rangle$  (left) and  $\langle \mathcal{J}_{ij} \rangle$  in units of the standard deviation (right) for each pair of spins vs.  $c_{ij}$ . For correlated spins, higher order coupling terms add up, on average, to strong pairwise couplings. Negative  $\mathcal{J}_{ij}$ 's typically correspond to positive correlations (and vice versa), as expected.

above  $\mathcal{D}^{(2)} \sim 0.3$ , and it was much lower for densely coupled networks, with lower entropy per spin. We traced the emerging pairwise sufficiency to formation of coherent clusters of spins, largely decoupled among themselves, resulting in a multitude of dependent attractors for the system. This is not a perturbative effect and is a new proposal for explaining pairwise sufficiency. Such collective behavior introduces substantial redundancy, and would allow error correction. However, this error correction is of a very different nature compared to, for example, the Hopfield network [31].

Does the mechanism presented here explain the pairwise sufficiency in any *real* biological system? This is unclear since our analysis was limited to specific simulated networks, which may or may not be good models of real biology. Specifically, the network in the original paper that observed the pairwise sufficiency [11] had much smaller entropy per spin (neurons rarely fired), and correlations among spins rarely exceed 0.2. In contrast, while 3-spin networks in our simulations had smaller  $\mathcal{S}$  and smaller spin-spin correlations than their 4-spin counterparts, these numbers were still larger than those in the experiments. At the same time, pairwise MaxEnt models do not fit experimental data perfectly (certainly worse than some of our nearly perfect fits) [15]. It may be that some structural features of real systems allow them to operate at higher  $\mathcal{D}^{(2)}$  for smaller  $\mathcal{S}$  compared to the simple models we investigated here — and exploring a wider class of networks for signatures of behaviors that we observed would be the next step. This is especially important since large coherent deviations from the most probable state into  $10 \dots 100$  metastable states seems to be a crucial feature of many experimental systems (such as bursts of neural activity in the retina that predominantly stays quiet [15]). Such metastable states far away from the ground state at least resemble the models that we studied. In addition, MaxEnt models in other fields may have very different properties compared to those in neuroscience, including different typical entropies and correlation strengths. Therefore, we hope that our models and their generalizations will

be able inform interpretation of experimental data, even if they do not match the experiments in some important properties.

With (approximate) pairwise sufficiency seen in many collective biological phenomena, it is important to ask why these systems operate in the regime that allows it to hold. Indeed, within our model, the pairwise sufficiency is not generic: low  $\mathcal{D}^{(2)}$  happens only for small  $\mathcal{S}$ , and preferentially when the strength of the interactions is high,  $\alpha = Ms/N \gg 1$  (cf. Fig. 2). The need for redundancy and error correction is a potential explanation – but there is no obvious reason why the redundancy must result in the pairwise sufficiency (indeed, simple parity-based codes probably do not). Taking the improvement in  $\mathcal{D}^{(2)}$  with the increase in  $\alpha$  seriously, we propose a different explanation (a similar argument was first suggested in Ref. [32]).

One can view evolution as trying to satisfy a growing list of constraints imposed upon a biological network by its interactions with the environment. These constraints can include efficient information processing, low energy consumption, robustness to perturbations, fitting within a certain physical size, responding quickly enough so that actions are relevant in the changing world, etc. Some of these global constraints may be equivalent to a large number of local constraints. For example, efficient information transmission in the visual system typically includes removal of redundancy present in the natural stimuli [33], which is equivalent to a multitude of constraints on activities of nearby neurons. When constraints are added, fewer and fewer states of the network remain accessible. Importantly, at least for certain abstract constraint satisfaction problems [30,34], before there are no more states left, the accessible states organize themselves in a handful of small, well-separated groups. Whether these states are uncorrelated, or consist of collective flipping of clusters of spins, they can be well represented by pairwise MaxEnt models. (In the former case, such MaxEnt model would have a Hopfield network structure [31]; in the latter, pairwise interactions would determine cluster assignments.) Therefore, it can be that the pairwise sufficiency is a signature of a biological network nearing the unsatisfiability threshold, being pushed towards it by evolution. Exploring landscapes of satisfiability problems with more realistic ensembles of constraints (or interactions) and comparing them to the landscapes observed in experiments would address this hypothesis.

**Acknowledgements** We thank Aly Pesic and Daniel Holz, who helped during the early stages of this project, and Arthur Lander and Chris Myers, who suggested a possible link to evolution. We are grateful to the Emory College Emerson Center for Scientific Computing and its funders for the help with numerical simulations. The authors were partially supported by the James S. McDonnell foundation Complex Systems award, by the Human Frontiers Science Program, and by the National Science Foundation.

## References

1. ET Jaynes. Information theory and statistical mechanics. *Phys Rev*, 106:620, 1957.
2. E Schneidman, S Still, MJ Berry, and W Bialek. Network information and connected correlations. *Phys Rev Lett*, 91:238701, 2003.

3. D Ackley, G Hinton, and T Sejnowski. A Learning Algorithm for Boltzmann Machines. *Cogn Sci*, 9:147, 1985.
4. T Broderick, M Dudik, G Tkacik, RE Schapire, and W Bialek. Faster solutions of the inverse pairwise Ising problem. *arXiv.org/0712.2437v2*, 2007.
5. M Weigt, R A White, H Szuromant, J A Hoch, and T Hwa. Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci (USA)*, 106:67, 2009.
6. M Mézard and T Mora. Constraint satisfaction problems and neural networks: A statistical physics perspective. *J Physiol (Paris)*, 103:107, 2009.
7. V Sessak and R Monasson. Small-correlation expansions for the inverse Ising problem. *J Phys A*, 42:055001, 2009.
8. S Cocco and R Monasson. Adaptive Cluster Expansion for Inferring Boltzmann Machines with Noisy Data. *Phys Rev Lett*, 106:090601, 2011.
9. HC Nguyen and J Berg. Mean-field theory for the inverse Ising problem at low temperatures. *Phys Rev Lett*, 109:050602, 2012.
10. G Tkacik and W Bialek. Information processing in living systems. *arXiv.org/1412.8752v1*, 2014.
11. E Schneidman, MJ Berry, R Segev, and W Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440:1007, 2006.
12. A Tang, D Jackson, J Hobbs, W Chen, J L Smith, H Patel, A Prieto, D Petrusca, M I Grivich, A Sher, P Hottowy, W Dabrowski, A M Litke, and J M Beggs. A Maximum Entropy Model Applied to Spatial and Temporal Correlations from Cortical Networks In Vitro. *J Neurosci*, 28:505, 2008.
13. I Ohiorhenuan, F Mechler, K Purpura, A Schmid, Q Hu, and J Victor. Sparse coding and high-order correlations in fine-scale cortical networks. *Nature*, 466:617, 2010.
14. G Field, J Gauthier, A Sher, M Greschner, T Machado, L Jepson, J Shlens, D Gunning, K Mathieson, W Dabrowski, L Paninski, A Litke, and EJ Chichilnisky. Functional connectivity in the retina at the resolution of photoreceptors. *Nature*, 467:673, 2010.
15. G Tkacik, O Marre, D Amodei, E Schneidman, W Bialek, and MJ Berry. Searching for collective behavior in a large network of sensory neurons. *PLoS Comp Bio*, 10:e1003408, 2014.
16. Bethge M and Berens P. Near-maximum entropy models for binary neural representations of natural images. In Platt J, Koller D, Singer Y, and Roweis S, editors, *Adv Neural Inf Proc Syst 20*, page 97. MIT Press, 2008.
17. N Halabi, O Rivoire, S Leibler, and R Ranganathan. Protein sectors: evolutionary units of three-dimensional structure. *Cell*, 138:774, 2009.
18. T Mora, A Walczak, W Bialek, and C Callan. Maximum entropy models for antibody diversity. *Proc Natl Acad Sci (USA)*, 107:5405, 2010.
19. W Bialek, A Cavagna, I Giardina, T Mora, E Silvestri, M Viale, and A Walczak. Statistical mechanics for natural flocks of birds. *Proc Natl Acad Sci (USA)*, 109:4786, 2012.
20. AA Margolin, K Wang, A Califano, and I Nemenman. Multivariate dependence and genetic networks inference. *IET Syst Biol*, 4:428, 2010.
21. J Otwinowski and I Nemenman. Genotype to Phenotype Mapping and the Fitness Landscape of the *E. coli lac* Promoter. *PLoS One*, 8:e61570, 2013.
22. T Mora and W Bialek. Are biological systems poised at criticality? *J Stat Phys*, 144:268, 2011.
23. DJ Schwab, I Nemenman, and P Mehta. Zipf’s law and criticality in multivariate data without fine-tuning. *Phys Rev Lett*, 113:068102, 2014.
24. G Tkacik, E Schneidman, M Berry, and W Bialek. Ising models for networks of real neurons. *arXiv.org/q-bio/0611072v1*, 2006.
25. Y Roudi, S Nirenberg, and P Latham. Pairwise maximum entropy models for studying large biological systems: when they can work and when they can’t. *PLoS Comp Biol*, 5:e1000380, 2009.
26. M Mézard, F Ricci-Tersenghi, and R Zecchina. Two solutions to diluted p-spin models and XORSAT problems. *J Stat Phys*, 111:505, 2003.

- 
27. T Kirkpatrick and D Thirumalai. p-spin-interaction spin-glass models: Connections with the structural glass problem. *Phys Rev B*, 36:5388, 1987.
  28. WE Deming and FF Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals Math Stat*, 11:427, 1940.
  29. M Mézard, G Parisi, and R Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297:812, 2002.
  30. F Krzakala, A Montanari, F Ricci-Tersenghi, G Semerjian, and L Zdeborova. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc Natl Acad Sci (USA)*, 104:10318, 2007.
  31. JJ Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci (USA)*, 79:2554, 1982.
  32. A Lander. Beyond q-bio? The closing talk at The Sixth q-bio Conference, Santa Fe, NM, 2012.
  33. HB Barlow. Possible principles underlying the transformation of sensory messages. In W Rosenblith, editor, *Sensory communication*, page 217. MIT Press, Cambridge, MA, 1961.
  34. CR Myers. Satisfiability, sequence niches and molecular codes in cellular signalling. *IET Syst Biol*, 2:304, 2008.